

11-01-00

A

Box Patent Application
ASSISTANT COMMISSIONER FOR PATENTS
Washington D.C., 20231

FORM PTO-1082
Case Docket No.: 81674.026 4196
Date: October 31, 2000
Express Mail Label No.: EL 669 015 584 US

Dear Sir:

Transmitted herewith for filing is the patent application of
Inventor(s): Robert G. Gally of Beaverton, Oregon; Eric W. Multanen of Portland, Oregon;
and Per Flemming Hanson of Soeborg, Denmark
For: DISTRIBUTED SWITCH/ROUTER SILICON ENGINE

Enclosed are:

- ☒ 5 Sheets of formal drawings
☒ An unsigned Declaration and Power of Attorney.

CALCULATION OF FEES						
ITEM		NO. OF CLAIMS FILED MINUS BASE*	NO. OF CLAIMS OVER BASE	X SM/LG ENTITY FEE	\$ AMOUNT	FEE
A	TOTAL CLAIMS FEE	23 -20*=	3	x \$9 or x \$18	\$ 54	
B	INDEPENDENT CLAIMS FEE**	3 -3*=	0	x\$40 or x 80	\$0	
C	SUBTOTAL – ADDITIONAL CLAIMS FEE (ADD FINAL COLUMN IN LINES A + B)					\$54
D	MULTIPLE-DEPENDENT CLAIMS FEE			SMALL ENTITY FEE = \$130 LARGE ENTITY FEE = \$260		\$0
E	BASIC FEE*			SMALL ENTITY FEE = \$355 LARGE ENTITY FEE = \$710		\$710
F	TOTAL FILING FEE (ADD TOTALS FOR LINES C, D, AND E)					\$764
G	ASSIGNMENT RECORDING FEE				\$	\$0
	**LIST INDEPENDENT CLAIMS 1, 9 and 18					

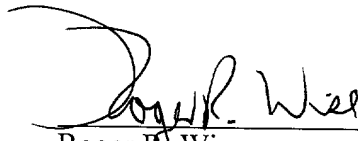
- _____ Please charge my Deposit Account No. \$ _____ A copy of this sheet is
_____ the amount of _____ enclosed.
☒ A check in the amount of \$764 to cover the filing fee is
enclosed.
_____ A check in the amount of \$ _____ to cover Assignment
Recordation fee is enclosed.
☒ The Commissioner is hereby authorized to charge payment of the following fees
associated with this communication or credit any overpayment to Deposit Account No.
03-3975. A copy of this sheet is enclosed.
☒ Any filing fees under 37 CFR 1.16 for the presentation of extra claims.
☒ Any patent application processing fees under 37 CFR 1.17.

_____ The Commissioner is hereby authorized to charge payment of the following fees during the pendency of this application or credit any overpayment to Deposit Account No. 03-3975.

- _____ Any patent application processing fees under 37 CFR 1.17.
- _____ The issue fee set in 37 CFR 1.18 at or before mailing of the Notice of Allowance, pursuant to 37 CFR 1.311(b).
- _____ Any filing fees under 37 CFR 1.16 for presentation of extra claims.

Respectfully submitted,

Dated: October 31, 2000



Roger R. Wise
Reg. No. 31,204

PILLSBURY MADISON & SUTRO LLP
725 South Figueroa Street, Suite 1200
Los Angeles, CA 90017-5443
Telephone: (213) 488-7100
Facsimile: (213) 629-1033

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

GALLY, et al.

Group No.: NOT ASSIGNED

Serial No.: NOT ASSIGNED

Examiner: NOT ASSIGNED

Filed: October 31, 2000

For: DISTRIBUTED SWITCH/ROUTER SILICON
ENGINE

CERTIFICATE OF MAILING VIA U.S. EXPRESS MAIL

"Express Mail" Mailing Label No. EL 669 015 584 US

Date of Deposit: October 31, 2000

Box Patent Application
Assistant Commissioner for Patents
Washington, D.C. 20231

Dear Sir:

I hereby certify that

X Letter of transmittal
X Check in the amount of \$764 as filing fee.
X Patent application (23 pages of specification; 23 claims; 1 pages of abstract)
X 5 sheets of formal drawings
X An unsigned Declaration
X Return postcard

are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service with sufficient postage under 37 CFR 1.10 on the date indicated above and are addressed to:

Box Patent Application
Assistant Commissioner for Patents
Washington, D.C. 20231.

October 31, 2000
Date of Deposit

Cynthia Belton
Cynthia Belton

APPLICATION FOR
UNITED STATES PATENT
IN THE NAME OF

Robert G. Gally

and

Eric W. Multanen

and

Per Flemming Hanson

for

DISTRIBUTED SWITCH/ROUTER SILICON ENGINE

prepared by:
PILLSBURY MADISON & SUTRO LLP
1100 New York Avenue, N.W.
Ninth Floor, East Tower
Washington, D.C. 20005-7100
(213) 488-7100
Attorney Docket No. 81674-264196
Client Reference No. P7781

Express Mail No.: EL 669 015 584 US

DISTRIBUTED SWITCH/ROUTER SILICON ENGINE

BACKGROUND OF THE INVENTION

1. Field of the Invention:

5 The present invention relates to the field of network communications. More particularly, the present invention relates to systems and methods for providing a remote switching engine to monitor and control network traffic, wherein appended word source address port mapping is utilized.

2. Related Art:

10 Computer networks in business enterprises, such as a local area network (LAN), wide area network (WAN) or other Ethernet-based systems, facilitate communication among computer workstations. The pressure on these networks is steadily increasing. More and more users are demanding more information and faster speed from increasingly distributed locations. At the same time, demanding new applications and
15 excessive Internet use are not only changing bandwidth requirements, they are also altering traditional traffic patterns.

20 When LAN networks were first introduced in the 1980's, a physical limit was quickly reached because of the LAN cable limitations. LAN bridges were introduced to solve this problem, tying these cables together to form larger networks. The bridge allowed the transparent passing of packets between LAN segments. Moreover, these bridges could also eavesdrop on the packets and learn which media access control (MAC) addresses were on each LAN segment. This allowed them to keep unicast traffic on the appropriate LAN segment. To utilize the bridges, MAC level broadcasts

were required. Broadcasts not only used network bandwidth, but they also used processing power on every host system to which the broadcast was being passed. The processor on the host system had to analyze every broadcast packet up through the network layer to see if the packet was addressed to it. Eventually, MAC level

5 broadcasts became an intolerably large percent of the network traffic. To solve this problem, routers were introduced to segment the network into separate domains.

At the router boundary, all broadcasts were intercepted and the router would decide which LANs on which the broadcast would be propagated. To achieve this, the router would look into level 3 headers and force a network to be segmented into

10 network level broadcast domains. Although this solved the problem of excessive broadcasts within the network, it introduced an expensive device that would add latency, limit throughput and increase complexity of the network. To limit the throughput loss across a router, users were forced into topologies where servers and clients needed to remain within the same broadcast domain. Therefore, switches were then

15 introduced to allow the creation of Virtual Local Area Networks (VLAN), allowing users to segment their networks without the high costs of routers or low port count of bridges. The first generation switches forwarded packets through the VLAN without examining the packet validity until after the packet had been forwarded. These switches did not prevent the occurrence of unnecessary and excessive traffic across the VLAN, which

20 slowed down the network and required each end node and computer connected to the network to receive and analyze those packets. This led to the overall loss of network bandwidth. To solve this problem, second-generation switches were created.

The second generation switches implement broadcast isolation and level 3

network switching at the switch level through end-to-end learning sequences, or learning hits. The second-generation switch comprises a switching application specific integrated circuit (ASIC) and a central processing unit (CPU) connected to a plurality of ports. The switching ASIC has a database which enables it to look up addresses that it has previously obtained and to forward frames to the addresses. When frames are to be sent through a second-generation switch, or a number of them, the switch(es) has to become aware of the location of the sender and the receiver of the frames. That is, the switch(es) has to learn ports with which source addresses and destination addresses of the frames are associated and update the information into the database.

FIG. 1 shows normal control frame paths of a prior art system in which switching ASICs learn the ports where the sender and the receiver reside. Three stacked switches 10, 20, 30 are illustrated in FIG. 1. Each of these switches includes a local CPU and a switching ASIC. For example, the switch 10 includes a local CPU 12 and a switching ASIC 15. In a normal frame control path, such as control paths 13, 23, 33, frames received by the switch 10 with unknown addresses are sent to the local CPU 12 through a PCI bus for the required learning. This introduces the requirement of having a CPU in every platform containing a switch. Overheads, such as the PCI bus, memory, flash, etc. are also present. Together, they increase costs to a system having many of these platforms. In addition, with different local CPUs monitoring and managing network traffic separately, a single point of management is not achieved. Therefore, there is a need for a system and method to provide a system that eliminates the need for having a CPU in every platform while allowing a single logical platform that facilitates a single point of management.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows normal control frame paths of a prior art system;

Figure 2 shows a remote control frame path according to an embodiment of the present invention;

5 Figure 3 illustrates a frame transmitted in the remote control frame path of FIG. 2;

Figure 4 illustrates processes for providing remotely controlled frames according to an embodiment of the present invention; and

Figure 5 illustrates processes for providing source address port mapping in a frame according to an embodiment of the present invention.

10

11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

DETAILED DESCRIPTION

Embodiments of the present invention are directed to systems and methods for providing a remote switching processing device to monitor and control network traffic, wherein appended word source address port mapping is utilized. In one embodiment, the system preferably includes a number of distributed switching systems connected together in a network. In FIG. 2, three switching systems 100, 200, 300 are illustrated as an example. The switching systems 100, 200, 300 may, for example, be stacked Ethernet switches that generally function as a single large switch. At least one of the switching systems includes a remote switching processing device 110 that is utilized to monitor and control network traffic through the switching systems 100, 200, 300. Each of the switching systems 100, 200, 300 includes a switching chip or module for high-speed packet switching. Each of the switching chips 120, 220, 320 within the switching systems 100, 200, 300 is connected to a number of network ports that interconnect the switching systems 100, 200, 300 and hosts in the network. For example, the switching chip 120 is shown to be connected to three network ports, with stack port 131 connecting the switching system 100 and the switching system 200, and stack port 133 connecting the switching system 100 and the switching system 300.

As configured in FIG. 2, the switching system 100 that contains the remote processing device 110 may be referred to as a remote switching system. The switching systems 200, 300 containing only the switching chips 220, 320 may be referred to as distributed switching systems. The remote switching processing device 110 in the remote switching system 100 may, for example, be a central processing unit (CPU). The switching modules or chips 120, 220, 320 may, for example, be switching

Application Specific Integrated Circuits (ASICs). The switching ASICs 120, 220, 320 may, for example, perform level 4 switching functions, level 3 switching functions, level 2 switching functions, level 3 router functions, and/or level 4 router functions. Although the switching functions in this embodiment have been described using ASICs, the switching ASIC functions may be implemented in software using a high-speed CPU or by hardware configurations not dependent on ASICs.

In one embodiment, each of the switching ASICs 120, 220, 320 has a Media Access Control (MAC) address lookup database (not shown). A MAC function converts digital information, typically stored in memory in the form of a packet, into an actual Ethernet frame that can be transmitted on an Ethernet connection, or a frame received from the network connection which is stored in memory as a packet. The MAC address lookup database allows each of the switching ASICs 120, 220, 320 to look up MAC addresses that each has previously obtained and to forward packets or frames to the MAC addresses. For switching decisions that cannot be determined within the switching ASICs 220, 320 of the distributed switching systems 200, 300, the remote switching processing device 110 makes such switching decisions.

Conversations between devices on a network, such as the switching systems 100, 200, 300 can be thought of as a matter of requests and responses. For example, a sender may wish to send frames or packets to a receiver through a switching system(s). The sender and receiver may, for example, be a switch, router, device for switching and routing, or host connected to network ports. Before frames can be sent to the receiver through the switching system(s), the switching system(s) must learn the source address and destination address for the frames to be transmitted. The switching

system(s), and more specifically, a switching ASIC(s) within the switching system(s) has to become aware of the sender and the receiver, and vice versa. This is achieved by having the remote switching processing device 110 update the MAC address lookup database of the switching ASIC(s) and encoding an ingress switch number and incoming port number in an appended word of a frame transmitted to an egress switch. The MAC address lookup database of the switching ASIC(s) is also referred to as a switch silicon forwarding database.

In a scenario where a sender residing on port 233 wishes to send frames to a receiver through the switching system 200, the first frame, or a portion of the frames, is first transmitted from the sender to the switching system 200 through port 233. As the frame enters port 233, it is received by the switching ASIC 220. The switching ASIC 220 extracts the source address of the frame and learns that the sender is on port 233. The switching ASIC 220 also extracts destination address of the frame and sends it to the MAC address lookup database. At this point, the destination address does not exist in the MAC address lookup database, and the switching ASIC 220 has to learn the destination address and with which port the destination address is associated. Since the frame is going to an unknown location, the frame is sent to all ports. At some point, the receiver is going to receive the frame and send a response back to the switching ASIC 220. When the switching ASIC 220 receives this response, the response will come back on a single port. The switching ASIC 200 extracts the source address of the response and sends it to the MAC address lookup database. Since this source address does not exist in the MAC address lookup database, the switching ASIC 220 forwards the response to the remote switching processing device 110 in the form of a response

frame. This is accomplished by using one of the Ethernet ports. Stack port 131 is used as an illustrative example in FIG. 2.

The response frame indicates to the remote switching processing device 110 that this source address of the response is unknown. The response frame is further
5 packaged by the switch ASIC 220 in a manner such that the remote switching processing device 110 would recognize the response frame to be a special frame for the remote switching processing device 110. The remote switching processing device 110 recognizes this special frame and determines that the special frame is not to be forwarded to another location. Instead, the remote switching processing device 110 is
10 to consume the response frame, process it, and respond to the switching ASIC 220 with a processing device directive. In other implementations, the frames may be required to be forwarded and not consumed by the switching processing device 110.

The processing device directive from the remote switching processing device 110 instructs the switching ASIC 220 to first put in its MAC address lookup database that the
15 address of the response resides on the port through which the response was received. An identifier is also included in the processing device directive to tell the switching ASIC 220 to consume the frame and not to forward it. Thus, the next time the switching ASIC 220 encounters a source or destination address that coincides with the address of the response, the switching ASIC 220 knows with which port the source or destination
20 address is associated. By the remote switching processing device 110 updating the MAC address lookup database of the switching system 200 with the source address of the sender and the destination address from the response of the receiver, the switching ASIC 220 becomes aware of the sender and the receiver, and vice versa.

In particular, a switching ASIC will forward the first frame of the flow to the remote switching processing device 110 when the switching ASIC does not find a forwarding entry in its MAC address lookup database. The remote switching processing device 110 learns the incoming port number and the Ethernet address of the source address and updates it in its MAC address lookup database. By using Ethernet ports to send learning frames to, and receiving learning frames from, switching ASICs, the remote switching processing device 110 also programs the outgoing port number and the Ethernet address of the destination address into the MAC address lookup database. The first frame is then routed on the port that has the destination node connected through it. Once the entries are created in the MAC address lookup table for the source and destination, all the packets belonging to the flow are routed in hardware at wire speed. In one embodiment, if the switching ASIC 220 is enabled to do IP or IPX routing, then it performs a packet validation step that checks to see if the frames are correctly formatted and eligible for routing. In other embodiments, packets belonging to protocols other than IP and IPX will be switched in hardware at wire speeds using the Layer 2 switching algorithm.

FIG. 3 illustrates a frame transmitted in the remote control frame path of FIG. 2, wherein appended word source address port mapping is utilized to map previously unknown source addresses to a specific distributed switch ASIC and port number. An appended word facility is used for data and control packets on stacking ports. In the appended word facility, ingress switches are allowed to specify set of egress switches for each packet. Intermediate switches and cross-bars do not need to do any address lookup and can switch based only on the appended word. When the frames reach the

egress switch, this switch does an address search to determine what set of local ports should transmit the packet. If the address search is unsuccessful, the egress switch and/or ingress switch must learn and associate the address being searched. The appended word source address port mapping facilitates this address search and learning of the address.

In the embodiment shown in FIG. 3, a frame from port 233 is being transmitted from switching ASIC 220 to switching ASIC 120 via stack port 131. The switching ASIC 220 and port 233 are referred to as an ingress switch engine and an incoming port, respectively. The switching ASIC 120 is referred to as an egress switch engine. The ingress switch engine number and incoming port number are first encoded in an appended word of a frame being transmitted to egress switch engine(s). In FIG. 3, the appended word of an exploded frame view shows the number for switching ASIC 220 and the number for port 233. In one implementation, the appended word is 32 bits and is inserted into an Ethernet frame. This appended word may be added, read, or removed on ports configured for appended word. The information--the switching ASIC 200 number and the port 233 number--is propagated in the packet header when the frame is forwarded to a processing device connected to the egress switch engine. In FIG. 3, the processing device is the remote switching processing device 110 in the remote switching system 100. In other embodiments, the frames may simply be sent to a distributed switching system similar to the distributed switching systems 200, 300. In that case, either the frames are further forwarded to the remote switching processing device 110 or to a local processing device, such as a local CPU, in distributed switching systems. This allows the egress switch engine(s) to map previously unknown source

addresses to a specific distributed switch ASIC and port. In this case, the specific distributed switch engine and port are switching ASIC 220 and port 233. With the ingress switch engine number and incoming port number, source address and destination address of a frame can be obtained. In the case of the destination address,
5 it will be the source address of a response frame from a receiver.

Each switching ASIC creates its own mapping of MAC addresses to egress port numbers based upon the frames it receives and with the help of the remote switching processing device 110 updates the MAC address lookup databases or the distributed switch ASIC forwarding databases. Unknown address frames are sent to the remote
10 switching processing device 110, which learns ingress switch engines and incoming port numbers and updates this information in the MAC address lookup database or distributed switch ASIC forwarding database of the distributed switching systems. This is accomplished by using Ethernet ports to send learning frames to, and receive learning frames from, switching ASICs. This mechanism allows autonomous forwarding
15 databases to be compiled independently by all distributed switching ASICs and switching systems in a multi-switching systems without a software protocol. The advantage of each switch ASIC creating its own forwarding database is that no distribution of learned information is required.

In order for the present invention to be operative, the remote switching
20 processing device 110 needs to be able to uniquely identify the originating switching ASICs, such as the switching ASIC 220, in order to send the response back to the originating switching ASICs. Various ways may be implemented to achieve this. In one implementation, a simple logic device on each distributed switch board of a distributed

switching system inserts a unique MAC address into the switching ASIC of the distributed switching system at initialization or boot time. This unique MAC address is programmed into a Read-Only-Memory (ROM) on the distributed switchboard during the manufacturing process. When a distributed switching system powers on, it repeatedly
5 broadcasts a frame with an appended word that indicates the unique MAC address of its switching ASIC and the fact that it is currently unmanaged. When the remote switching processing device 110 receives this frame, the remote switching processing device 110 associates a unique engine number with the received unique MAC address. The remote switching processing device 110 then transmits a CPU control frame with
10 appended word to the distributed switch system, directing the distributed switch ASIC to use to associated engine number in all subsequent frame appended words.

In one embodiment, learning frames are tagged as higher than normal traffic priority. This is necessary because these frames are used for managing traffic and needs to be resolved first before the actual transmitting of frames is to proceed. The
15 highest priority queue is needed to minimize frame loss. In one implementation, a queuing engine is provided in a switching system, preferably in the switching ASIC of the switching system. This includes both the enqueueing and dequeuing logic. Each switching ASIC is to support unique levels of priority queues, with the highest priority being assigned to frames that are used exclusive for managing traffic. For example,
20 frames for resolving the source and destination addresses and determining transmit ports need to be assigned with highest priority.

Several advantages are realized with the present invention. With a remote switching processing device, associations between MAC and network ports are learned

through the distributed switching ASIC forwarding unknown address frames to the remote switching processing device. These forwarded unknown address frames are forwarded to the remote switching processing device using Ethernet ports. Each forwarded unknown address frame has an appended word containing an ingress switch engine number and an incoming port number. The remote switching processing device then updates the forwarding database of the distributed switching ASIC with this information. By utilizing the remote switching processing device and the Ethernet ports to learn associations between MAC and network ports, a processing device, such as a local CPU, does not to be present on every platform or switching system. Only the switching system containing the remote switching processing device needs to have a processing device. This reduces costs dramatically. Moreover, processing devices, such as CPUs, come with substantial overhead. Illustrative examples of such overhead are PCI buses, memory, flashes, and a number of other devices. By eliminating the need for a processing device, the need for the corresponding overhead is also eliminated. In embodiments where local processing devices are provided to distributed switching systems to allow localized optimization of some local CPU functions, low end CPUs can be utilized because the local processing device does not need to be involved in monitoring or controlling network traffic. This also saves system costs.

According to an embodiment of the present invention, the remote switching processing device 110 is utilized to allow a more general operation of having net identifications (netIDs) to supplant local CPU queues. The netIDs contain the append word feature, which is used to cascade other devices using a switching ASIC as a switching matrix. The NetIDs also contain the source addresses and destination

addresses based mirror port information for global source and global destination address based mirroring. Frames which normally would go to a local switching processing device, such as a local CPU, are instead transmitted to the remote processing device 110 coupled to the switching ASIC 120 elsewhere in the stack of switches. In this case, the remote switching processing device 110 also needs to be able to uniquely identify an originating switching ASIC, so that the remote switching processing device 110 can respond to the originating switching ASIC. The frames also need to be tagged as higher than normal traffic priority. CPU queue number should also be preserved, e.g., having a unique netID per CPU queue.

Upon receiving these frames, the remote switching processing device 110 processes these frames. If necessary, the remote switching processing device 110 responds by transmitting netID appended frames to an originating switching ASIC and indicating the response as a "processing device directive." When these netID appended frames are received by the originating switching ASIC, these frames are processed just as if they were originated locally from a local switching processing device. In one implementation, secure ports are provided between different switching systems, such as the switching system 100 and switching system 200, and only processing device directives from secure ports are accepted. A secure port may, for example, be the stack port between the switching systems 100 and 200. In other embodiments, security ports are implemented using security protocols.

In one embodiment, each of the distributed switching systems 200, 300 are provided with a local processing device, such as a local CPU. The local processing device may be a low end processing device as compared to the remote switching

processing device 110. This is because the local processing device does not need to be involved in monitoring and managing network traffic, e.g., with packet transfers to and from the switching ASICs. With local processing devices in the distributed switching systems 200, 300, not all processing device queues need to be sent to the remote switching processing device 110. This allows localized optimization of some processing device functions and allows the remote switching processing device 110 to send frames to the local processing devices. With low end processing devices, cost optimized distributed switching systems are achieved. The advantage of this implementation is a streamlined control flow of externally interconnected switching ASICs that can be managed as a single logic platform. For example, the configuration may be used to facilitate Single Point of Management (SPOM) in stackable switching router products, including 10/100 Mb 24 port stackable Ethernet switches, 10/100/1000 Mb 8 port stackable Ethernet routing switch, 10/100 Mb 24 port stackable Ethernet switch with stacking crossbar, and 10/100/1000 24 port stackable Ethernet routing switch. The SPOM feature gives a device manager the ability to manage a whole stack as one device with one IP address and gives a user the look and feel that a stack of switches is managed as a single device.

Figure 4 illustrates processes for providing remotely controlled frames to monitor and control network traffic according to an embodiment of the present invention. In one embodiment, the system includes a number of local switching devices and a remote switching processing device. In block P400, a frame destined for a destination port is received by a local switching device from a source port. One of the elements included in the frame is a destination address of the destination port. In block P410, the

destination address of the destination is analyzed. It is determined if the destination address of the frame is known in a Media Access Control (MAC) address database. In block P415, if the destination address is known in the MAC address database, the frame is forwarded to the destination port corresponding to the destination address. In block P420, the destination address is not previously known in the MAC address database, and an unknown destination address frame is forwarded to all ports asking for a response. In block P430, when the receiving port receives the unknown destination address frame, the receiving port sends a response frame back to the local switching device, where the unknown destination address frame originated in this embodiment. In block P440, it is determined if the source address of the response frame is known in the MAC address database. In block P445, if the source address of the response frame is known, the frame is forwarded to the port corresponding to the source address of the response frame. In block P450, if the source address of the response frame is not known previously, the response frame is forwarded to the remote processing switching device. In block P460, based on the received response frames as well as associated addresses and ports, the remote switching processing device updates the MAC address database corresponding to the local switching device and the local switching device learns associations between MAC addresses and ports.

Figure 5 illustrates processes for providing appended word source address port mapping in a frame and allowing autonomous forwarding database to be compiled according to an embodiment of the present invention. In block P510, a frame is transmitted from an ingress switch engine to an egress switch engine. In block P520, an ingress switch engine number and an incoming port number are encoded in an

appended word of the frame. In one implementation, the numbers are encoded before the frame is transmitted. In other embodiments, the numbers are encoded during the transmission. The ingress switch engine number indicates a specific switching device from which the frame is being transmitted. The incoming port number indicates the port number of an incoming port from which the frame originated. In block P530, the encoded information is forwarded to a processing device of the egress switch engine. The processing device may, for example, be a CPU. In block P540, it is determined whether a source address of the frame is previously known. In block P550, when the source address is not previously known, the egress switch engine maps the unknown source address to the ingress switch engine number and the incoming port number.

While the description above refers to particular embodiments of the present invention, it will be understood that many modifications may be made without departing from the spirit thereof. For example, a switch/router ASIC that performs the functions of both conventional a switch and a router may be implemented in place of a switch ASIC that only performs the function of a conventional switch. Moreover, although the inventive concepts described herein utilize Ethernet protocols, these concepts are readily applicable to other types of networks. The accompanying claims are intended to cover such modifications as would fall within the true scope and spirit of the present invention. The presently disclosed embodiments are therefore to be considered in all respects as illustrative and not restrictive, the scope of the invention being indicated by the appended claims, rather than the foregoing description, and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced therein.

CLAIMS

What is claimed is:

1. A system comprising:

5 a plurality of ports for sending and receiving frames;

a local switching device that performs high-speed switching;

a local forwarding database corresponding and coupled to the local switching device, the database allowing the local switching device to look up a known address that has been previously obtained and forward the frames based on the known address;

10 and

a remote switching processing device that receives and processes frames, wherein the local switching device learns associations between Media Access Control (MAC) addresses and ports by having the local switching device forward unknown address frames to the remote switching processing device, and the remote switching processing device update the local forwarding database corresponding to the local switching device based on the forwarded frames.

2. The system of claim 1, wherein each of the forwarded frames includes an appended word, the appended word being encoded with an ingress switch engine number and an incoming port number, the ingress switch engine number indicating a specific local switching device, the incoming port number indicating the incoming port number of a port from which unknown address frames are being sent, the ingress switch engine and incoming port numbers being used to allow an egress switch engine

to map the unknown addresses to the ingress switch engine number and the incoming port number.

3. The system of claim 2, wherein the ingress switch engine is the local switching
5 device, and the incoming port is a port connected to the local switching device.

4. The system of claim 1, wherein the remote switching processing device updates
the local forwarding database by sending a processing device frame with an appended
word.

10 5. The system of claim 1, wherein the ports are Ethernet ports and the frames being
sent, received, and forwarded are Ethernet frames.

15 6. The system of claim 1, further comprising logic residing on the local switching
device to insert a unique MAC address into the local switching device, wherein the
unique MAC address is provided to allow the remote switching processing device to
uniquely identify the local switching device.

20 7. The system of claim 1, wherein the unknown address frames are tagged as
higher than normal traffic priority.

8. The system of claim 1, further comprising a local processing device coupled to
the local switching device, the local processing device allowing localized optimization of

local functions and allowing the remote switching processing device to send frames to the local processing device.

9. A method for monitoring and controlling network traffic in a system having a local

5 switching device and a remote switching processing device, the method comprising:

receiving a frame from a source port, the frame being destined for a destination port indicated by a destination address of said frame;

determining if the destination address of said frame are known in a Media Access Control (MAC) address database;

10 forwarding the frame to the destination port when the destination address is known in the MAC address database;

sending an unknown destination address frame to all ports asking for a response frame when the destination address is unknown in the MAC address database;

receiving said response frame from a receiving port;

15 determining if a source address of the response frame is known in the MAC address database; and

forwarding the response frame to the remote processing switching device,

wherein the local switching device learns associations between MAC addresses and ports by having the remote switching processing device update the MAC address

20 database corresponding to the local switching device based on the received response frames.

10. The method of claim 9, wherein each of the received frames from a source port and the response frame from a destination port includes an appended word, the appended word being encoded with an ingress switch engine number and an incoming port number, the ingress switch engine number indicating a specific local switching device, the incoming port number indicating the incoming port number of a port from which unknown address frames are being sent, the ingress switch engine and incoming port numbers being used to allow an egress switch engine to map the unknown addresses to the ingress switch engine number and the incoming port number.

11. The method of claim 10, wherein the ingress switch engine is the local switching device, and the incoming port is a port connected to the local switching device.

12. The method of claim 9, wherein the MAC address database is updated by the remote switching processing device sending a processing device frame with an appended word.

13. The method of claim 9, wherein the source and destination ports are Ethernet ports and the frames are Ethernet frames.

14. The method of claim 9, further comprising providing logic residing on the local switching device to insert a unique MAC address into the local switching device, wherein the unique MAC address is provided to allow the remote switching processing device to uniquely identify the local switching device.

15. The method of claim 9, wherein the unknown destination address frames are tagged as higher than normal traffic priority.

5 16. The method of claim 9, further comprising providing a local processing device coupled to the local switching device, the local processing device allowing localized optimization of local functions and providing the remote switching processing device to send frames to the local processing device.

10 17. The method of claim 9, further comprising:
determining if a source address of the frame from the source port is known in the MAC address database; and
updating the source address of said frame in the MAC address database if the source address is not known.

15 18. The method of compiling autonomous forwarding databases using appended word source address port mapping:
transmitting a frame from an ingress switch engine to an egress switch engine;
encoding an ingress switch engine number and an incoming port number in an
20 appended word of the frame, the ingress switch engine number indicating a specific switching device from which the frame is being transmitted, the incoming port number indicating the port number of an incoming port from which the frame originated;

forwarding encoded information to a processing device of the egress switch engine;

determining whether a source address of the frame is previously known; and

allowing the egress switch engine to map an unknown source address to the

5 ingress switch engine number and the incoming port number.

19. The method of claim 18, wherein the appended word resides in the packet header of the frame.

10 20. The method of claim 18, wherein the ingress switch engine is a local switching device, and the incoming port is a port connected to the local switching device.

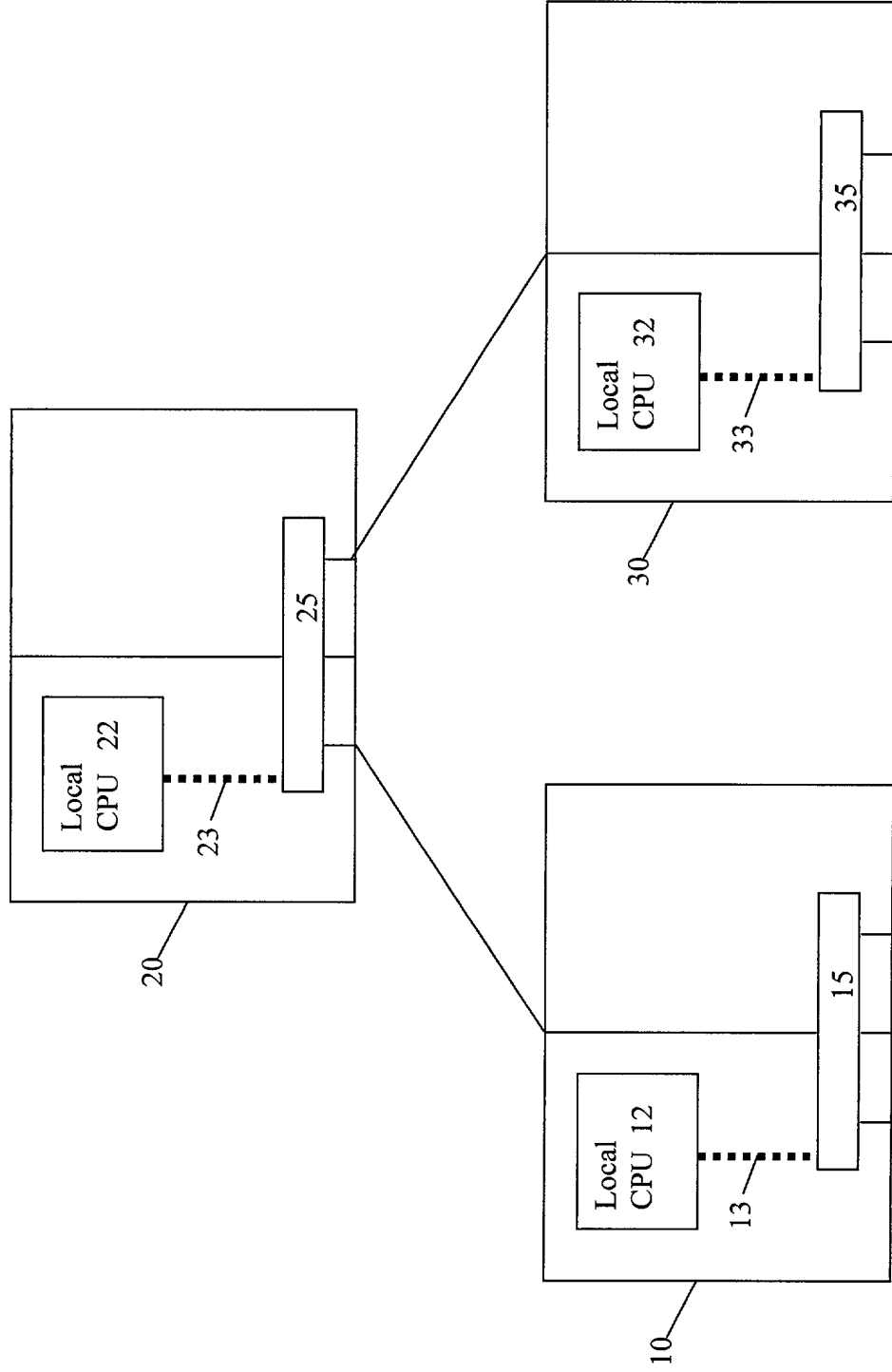
21. The method of claim 20, wherein the egress switch engine is also a local switching device.

15 22. The method of claim 20, wherein the egress switch engine is a remote switching device containing a remote switching processing device.

20 23. The method of claim 18, wherein the ingress switch engine is a remote switching device containing a remote switching processing device, the incoming port is a port connected to the remote switching device, and the egress switch engine is a local switching device.

ABSTRACT

A method and system for providing a remote switching engine to monitor and control network traffic and utilizing appended word source address port mapping is provided. The system comprises a number of ports, at least one local switching device, at least one local forwarding database, and a remote switching processing device. The ports are provided for sending and receiving frames. The local switching device performs high-speed switching. The local forwarding database corresponds and couples to the local switching device, the database allowing the local switching device to look up a known address that has been previously obtained and forward the frames based on the known address. The remote switching processing device receives and processes frames from the local switching device(s). The local switching device(s) learns associations between Media Access Control (MAC) addresses and ports by having the local switching device forward unknown address frames to the remote switching processing device. The remote switching processing device update the local forwarding database corresponding to the local switching device based on the forwarded frames, utilizing the appended word source address mapping. The appended word of a frame transmitted from one switching device to another switching device has encoded within an ingress engine number and port number.



..... Normal Control Frame Path

Figure 1
(Prior Art)

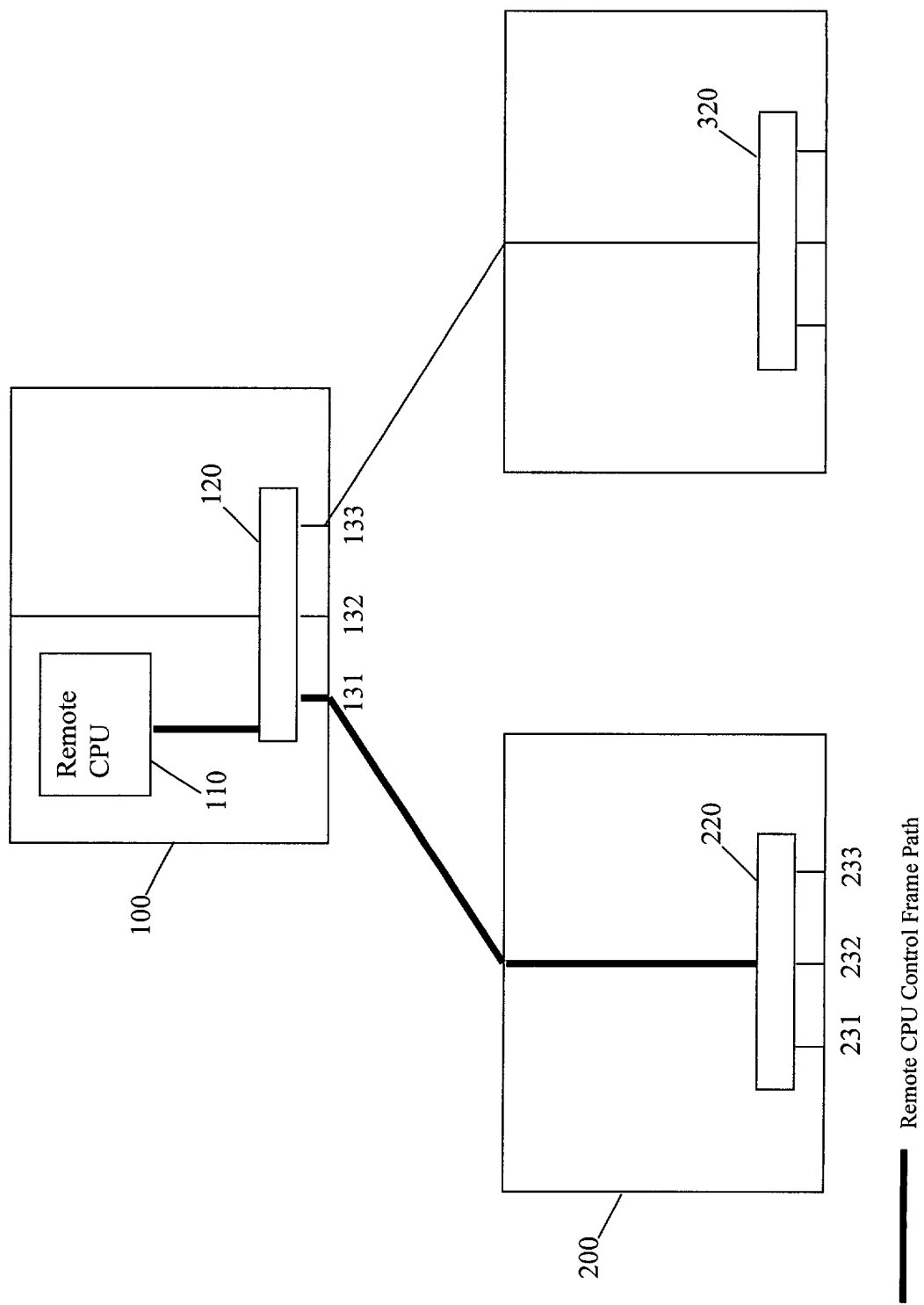
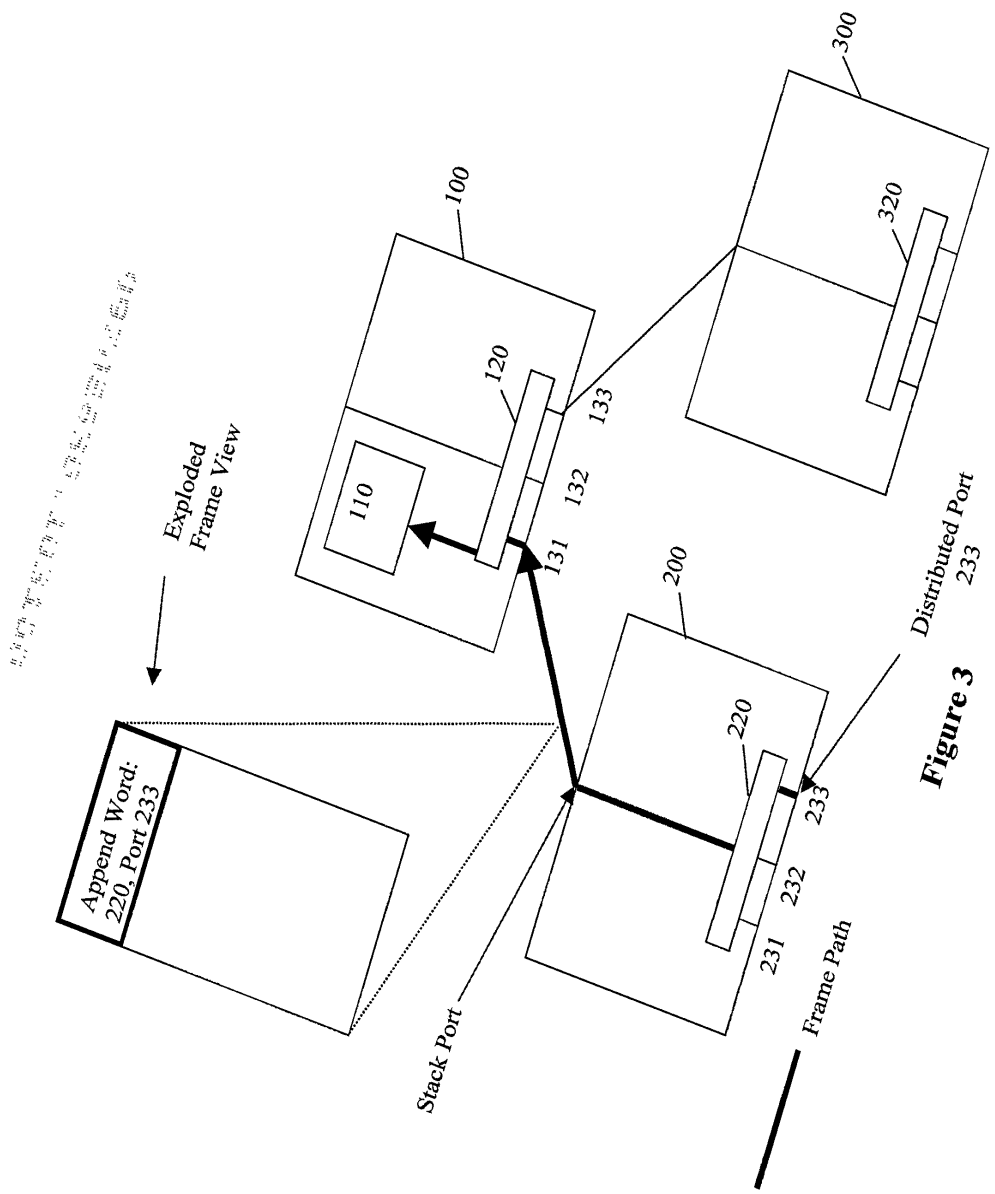


Figure 2



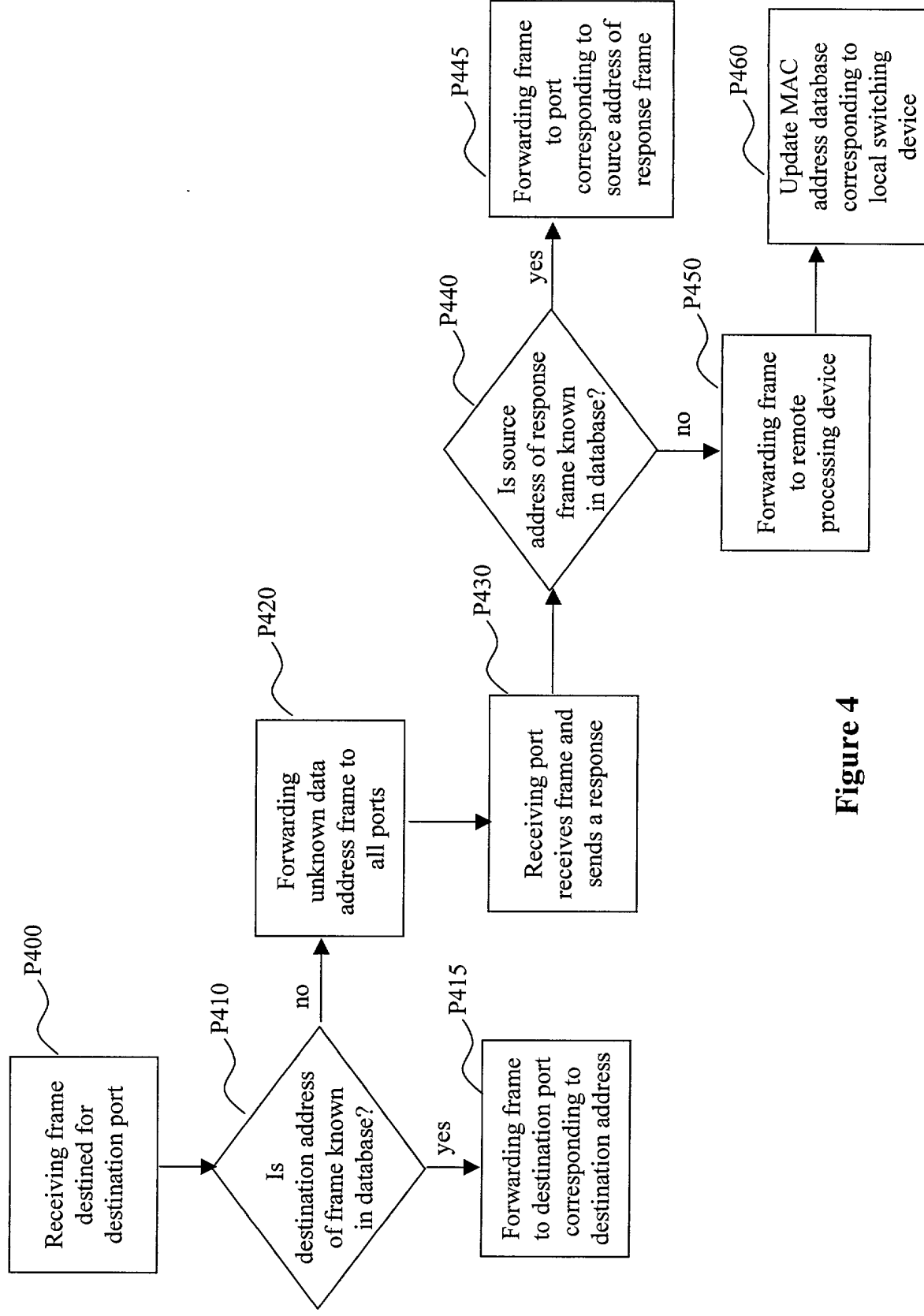


Figure 4

FOR UTILITY/DESIGN
CIP/PCT NATIONAL/PLANT
ORIGINAL/SUBSTITUTE/SUPPLEMENTAL
DECLARATIONS

RULE 63 (37 C.F.R. 1.63)
DECLARATION AND POWER OF ATTORNEY
FOR PATENT APPLICATION
IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

PM & S
FORM

As a below named inventor, I hereby declare that my residence, post office address and citizenship are as stated below next to my name, and I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the INVENTION ENTITLED
DISTRIBUTED SWITCH/ROUTER SILICON ENGINE

the specification of which (CHECK applicable BOX(ES))

X
A ☒ is attached hereto.

BOX(ES) → B. ☐ was filed on _____ as U.S. Application No. _____ /

→ C. ☐ was filed as PCT International Application No. PCT/ _____ / _____ on _____

and (if applicable to U.S. or PCT application) was amended on _____

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose all information known to me to be material to patentability as defined in 37 C.F.R. 1.56. Except as noted below, I hereby claim foreign priority benefits under 35 U.S.C. 119(a)-(d) or 365(b) of any foreign application(s) for patent or inventor's certificate, or 365(a) of any PCT International Application which designated at least one other country than the United States, listed below and have also identified below any foreign application for patent or inventor's certificate, or PCT International Application, filed by me or my assignee disclosing the subject matter claimed in this application and having a filing date (1) before that of the application on which priority is claimed, or (2) if no priority claimed, before the filing date of this application

PRIOR FOREIGN APPLICATION(S)

Number	Country	Day/MONTH/Year Filed	Date first Laid-open or Published	Date Patented or Granted	Priority NOT Claimed
--------	---------	----------------------	-----------------------------------	--------------------------	----------------------

If more prior foreign applications, X box at bottom and continue on attached page.

Except as noted below, I hereby claim domestic priority benefit under 35 U.S.C. 119(e) or 120 and/or 365(c) of the indicated United States applications listed below and PCT international applications listed above or below and, if this is a continuation-in-part (CIP) application, insofar as the subject matter disclosed and claimed in this application is in addition to that disclosed in such prior applications, I acknowledge the duty to disclose all information known to me to be material to patentability as defined in 37 C.F.R. 1.56 which became available between the filing date of each such prior application and the national or PCT international filing date of this application

PRIOR U.S. PROVISIONAL, NONPROVISIONAL AND/OR PCT APPLICATION(S)

Application No. (series code/serial no.)	Day/MONTH/Year Filed	Status	Priority NOT Claimed
--	----------------------	--------	----------------------

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

And I hereby appoint Pillsbury Madison & Sutro LLP, Intellectual Property Group, 1100 New York Avenue, N.W., Ninth Floor, East Tower, Washington, D.C. 20005-3918, telephone number (202) 861-3000 (to whom all communications are to be directed), and the below-named persons (of the same address) individually and collectively my attorneys to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith and with the resulting patent, and I hereby authorize them to delete names/numbers below of persons no longer with their firm and to act and rely on instructions from and communicate directly with the person/assignee/attorney/firm/ organization who/which first sends/sent this case to them and by whom/which I hereby declare that I have consented after full disclosure to be represented unless/until I instruct the above firm and/or a below attorney in writing to the contrary

Paul N. Kokulis	16773	Paul E. White, Jr.	32011	Stephen C. Glazier	31361	Adam R. Hess	41835
Raymond F. Lippitt	17519	Glenn J. Perry	28458	Ruth N. Morduch	31044	William P. Atkins	38821
G. Lloyd Knight	17698	Kendrew H. Colton	30368	Richard H. Zaitlen	27248	Paul L. Sharer	36004
Kevin E. Joyce	20508	G. Paul Edgell	24238	Roger R. Wise	31204		
George M. Sirilla	18221	Lynn E. Eccleston	35861	Jay M. Finkelstein	21082		
Donald J. Bird	25323	Timothy J. Klima	34852	Michael R. Dzwonczyk	36787		
Peter W. Gowdey	25872	David A. Jakopin	32995	W. Patrick Bengtsson	32456		
Dale S. Lazar	28872	Mark G. Paulson	30793	Jack S. Barufka	37087		

(1) INVENTOR'S SIGNATURE:

Date:

Robert	G.	GALLY
First	Middle Initial	Family Name
Residence	Beaverton	Oregon
City	State/Foreign Country	Country of Citizenship
Post Office Address	14820 SW Bonnie Brae Street, Beaverton, Oregon 97007	
(include Zip Code)	97007	

(2) INVENTOR'S SIGNATURE:

Date:

Eric	W.	MULTANEN
First	Middle Initial	Family Name
Residence	Portland	Oregon
City	State/Foreign Country	Country of Citizenship
Post Office Address	170 NW 114th Avenue, Portland, Oregon 97229	
(include Zip Code)	97226	

FOR ADDITIONAL INVENTORS, "X" box ☒ and proceed on the attached page to list each additional inventor.
☐ See additional foreign priorities on attached page (incorporated herein by reference).

Atty. Dkt. No. PM81674.026 4196
(M#)

DECLARATION AND POWER OF ATTORNEY

(continued)

ADDITIONAL INVENTORS:

(3) INVENTOR'S SIGNATURE:

Date:

Per	Flemming	HANSON
First	Middle Initial	Family Name
Residence	Soeborg	Denmark
City	State/Foreign Country	Country of Citizenship
Post Office Address	Maglegaards Alle 120, 2862 Soeborg, Denmark	
(include Zip Code)	2862	